EXHIBIT A

VirtualMiner Software Requirements Specification

1 Introduction

VirtualMiner is a software component that will enable users of various kinds of data to view, filter, query and mine their data easily. The target audience consists of "lay" users of data (as opposed to trained statisticians and data analysts). VirtualMiner consists of two parts, the Analyzer and the Viewer. Used in conjunction, the two components provide users with an easy and intuitive way of interacting with and discovering knowledge in their data.

In a multi-user scenario, such as an intranet/Internet, the Analyzer would typically be used by a data analyst/webmaster to prepare the raw data for analysis. Once this has been done, the end-user can then use the Viewer to view, filter and query the data, and view data mining results.

In a single-user scenario, the Analyzer and Viewer can be used by the same person to set up data for analysis, and view the results.

Both the Analyzer as well as the Viewer can be executed from within tools that developers as well as end-users are already familiar with, including web browsers and popular visual development tools and database and spreadsheet applications. Thus, there is no need for re-learning a complex user interface.

1.1 References

1. (AF Paper)

2 The VirtualMiner Analyzer

2.1 Introduction

The VirtualMiner Analyzer enables a designer to:

- Set up raw data to enable viewing, filtering and querying
- Set up one or more Data Mining analyses on the data (the Analyzer uses the Attribute Focusing algorithm for mining data [1])

2.2 Analyzer Execution

The VirtualMiner Analyzer should be a software component, capable of running within several environments, viz.

- Web browsers, such as Microsoft Internet Explorer and Netscape Navigator
- Visual Development tools that are capable of importing software components, such as Microsoft Visual C++, Microsoft Visual Basic, Microsoft Visual J++, Symantec Café, Borland C++, Sun Java WorkShop, Sun BeanBox, etc.

- Popular database utilities, such as Microsoft Access and Lotus Approach
- Popular spreadsheet utilities, such as Microsoft Excel, and Lotus 1-2-3

2.3 Inputs

The VirtualMiner Analyzer should be able to handle raw data in the following formats:

- ASCII Text Files
- Microsoft Excel and Lotus 1-2-3 Spreadsheets
- Local databases - Microsoft Access, Paradox, Dbase, Lotus Approach
- All RDBMS servers – Oracle, Sybase, Informix, SQLServer, Gupta, DB2
- Tables embedded in HTML files

2.4 Processing

The processing by the VirtualMiner Analyzer achieves the following purposes:

- Enables fast OLAP queries on the raw data
- Enables fast and easy Data Mining queries on the raw data

The Analyzer should read the raw data, and process it, using the following parameters provided by the user:

- Column(s) of interest for OLAP queries
- Data Mining Questions, each question consisting of :
  - Variable(s) of interest
  - Decision Variable
  - Numeric Attribute, if any
  - Cutoff

(The User Interface for collecting these parameters is described in Section 2.6)

2.5 Outputs

In order to achieve the aims of fast and easy querying on the raw data, it may be necessary for the Analyzer to convert the raw data into a format more amenable to fast querying. In such a situation, the processed data should be stored along with the raw data, and linked to it in some manner, so that when the user views the raw data, the results of processing are also immediately available (refer Section 3 for more details)

2.6 User Interface

The User Interface of the Analyzer consists of a wizard, which guides the designer through the process of setting up the parameters for processing the

data. At each step, the designed is shown sample results, using the inputs he / she has already provided. (If no inputs have been provided to the Analyzer, sample results shown should be based on "dummy" inputs).

On each screen, there will be a "Tutorial" button. Pressing this button would provide a short, animated tutorial on using the screen. The tutorial would provide help on using the controls on the screen, as well as suggest possible alternatives to the choices that the designer has made. This should be achieved in an easy-going, friendly and fun-to-use manner.

The purpose of the interface is to collect the following information:

- Type of input data (local database, ASCII text file, etc.)
- Name of file / database which contains data
- Name of table in which data are stored (required if input is from database / HTML file)
- Column(s) of interest for OLAP queries
- Data Mining Questions, each question consisting of the following :
  - Variable(s) of interest
  - Decision Variable
  - Numeric Attribute, if any
  - Cutoff

The User Interface consists of the following screens:

2.6.1 Screen 1 : Welcome Screen

This screen is invoked when the user starts the Analyzer. It displays a suitable welcome message, informing the designer of the process he / she is about to go through. The screen contains the following controls:

1. *Cancel Button*

Pressing this button will cause the Analyzer to exit

2. *Next Button*

Pressing this button will cause Screen 1 to close, to be replaced by Screen 2

2.6.2 Screen 2 : Data Source Screen

This screen is invoked when the user presses the *Next* button on Screen 1, or the *Previous* button on Screen 3. It contains the following controls:

1. *ASCII Text File Option*

The designer can select this option if the raw data are contained in an ASCII text file

2. *HTML File Option*
The designer can select this option if the raw data are contained in an HTML file.

3. *Local Database Option*
The designer can select this option if the raw data are contained in a table on a local database (Microsoft Access / Lotus Approach). Note that "local" in this context implies a "non-server" database. The database file itself could be located on a fileserver on a network. However, it is accessed as if it were a local file, and not through a database server, such as Sybase or Oracle.

4. *Database server Option*
The designer can select this option if the raw data are contained in a table on a database server.

5. *Spreadsheet Option*
The designer can select this option if the raw data are contained in a Microsoft Excel / Lotus 1-2-3 spreadsheet

6. *Next Button*
This button will be enabled if the data source type has been selected. Pressing this button will cause Screen 2 to close. One of the following actions is taken :

- If the designer selected the Database Server option, Screen 4 is displayed
- For all other options, Screen 3 is displayed

7. *Previous Button*
Pressing this button will cause Screen 2 to close, to be replaced by Screen 1

8. *Cancel Button*
Pressing this button will cause the Analyzer to exit

9. *Tutorial Button*
Pressing this button will cause the tutorial for this screen to start.

2.6.3 Screen 3 : File Select Dialog Box

This is the standard system File Select Dialog Box. It is displayed if the user has selected any option other than the Database server option on Screen 2. Using this screen, the designer should be able to select the file where the data are stored. (Depending on the selection in Screen 2, she would be able to select an ASCII text file, an MS Access or Lotus Approach database, an MS Excel or Lotus 1-2-3 spreadsheet, or an HTML file).

Once the designer selects a file, and closes this dialog box, one of the following actions is taken:

- If the designer selected an ASCII text file, Screen 5 is displayed
- If the designer selected a local database or HTML file, Screen 6 is displayed

- If the designer selected a spreadsheet, Screen 7 is displayed

If the designer closes this dialog box without selecting a file, control returns to Screen 2.

2.6.4 Screen 4 : ODBC Login Screen

This screen is the standard system ODBC Login screen for selecting a data source from a database server. It is displayed if the designer selects a Database server as the data source in Screen 2. The designer will be able to select any ODBC source for the data.

Upon successful login, this screen will close, to be replaced by Screen 7.

2.6.5 Screen 5 : ASCII Text File Import Options Screen

This screen is displayed if the designer has selected an ASCII text file as the data source. The designer can specify import options for the text file using this screen. This screen has the following controls:

1. *Field Separator drop-down list*
This list contains possible options for the field separator character. The program should guess the most appropriate character, which should be selected by default.

2. *Text Delimiter drop-down list*
This list contains possible options for the text delimiter character. The program should guess the most appropriate character, which should be selected by default.

3. *Next Button*
Pressing this button causes Screen 5 to close, and Screen 7 to be displayed

4. *Previous Button*
Pressing thus button causes Screen 5 to close, and Screen 2 to be displayed.

5. *Cancel Button*
Pressing this button causes the Analyzer to exit.

6. *Tutorial Button*
Pressing this button causes the tutorial for this screen to start

2.6.6 Screen 6 : Table Select Screen

This screen is displayed if the designer selected a local database, database server or HTML file as the data source. It enables the designer to select a table from which the data are imported.

This screen contains the following controls:

1. *Select Table List*
This is a list of the tables in the data source selected by the designer, of which she can select one.

2. *Next Button*
Pressing this button causes Screen 6 to close, to be replaced by Screen 7

3. *Previous Button*
Pressing this button causes Screen 6 to close, to be replaced by Screen 2.

4. *Cancel Button*
Pressing this button causes the Analyzer to exit

5. *Tutorial Button*
Pressing this button causes the tutorial for this screen to start

2.6.7 Screen 7 : Selecting an analysis template

This screen allows the designer to select a pre-defined analysis template (refer Section 2.6.15 for details on creating an analysis template). In this manner, the designer can apply the same analysis to different tables of the same type (for example, sales figures of different months)

This screen contains the following controls:

1. *Define a new Analysis template Option (default)*
Selecting this option implies that the designer does not wish to use a pre-defined template, but wants to define a new analysis template.

2. *Analyze using a saved template Option*
Selecting this option implies that the designer wishes to use a pre-defined template.

3. *Next Button*
Pressing this button causes Screen 7 to close. If the designer selected the first option (Define a new Analysis template), Screen 9 is displayed. If the designer selected Option 2 (Analyze using a saved template), Screen 8 is displayed.

4. *Previous Button*
Pressing this button causes Screen 7 to close, to be replaced by Screen 2.

5. *Cancel Button*
Pressing this button causes the Analyzer to exit.

6. *Tutorial Button*
Pressing this button causes the tutorial for this screen to start.

2.6.8 Screen 8 : Select Analysis Template Dialog Box

This is the standard system File Select Dialog Box, and allows the designer to select a saved Analysis Template file.

If the designer selects an analysis template file and closes this dialog box, Screen 16 is displayed.

If the designer closes this dialog box without selecting a file, control returns to Screen 7.

2.6.9 Screen 9 : Selecting columns for OLAP queries

This screen allows the designer to select columns that are important for end-user queries. This screen contains the following controls:

1. *Column List*
The columns in the selected table are displayed. The designer can select one or more columns from this table.

2. *Sample OLAP Query Textbox*
This control displays a sample query involving the column(s) selected by the designer. (The sample query displayed changes as and when the column(s) selected by the designer change(s))

3. *Next Button*
Pressing this button causes this screen to close, and Screen 10 to be displayed.

4. *Previous Button*
Pressing this button causes this screen to close, to be replaced by Screen 7.

5. *Cancel Button*
Pressing this button causes the Analyzer to exit.

6. *Tutorial Button*
Pressing this button causes the Tutorial for this screen to start.

2.6.10 Screen 10 : Selecting a Numeric Variable for Data Mining

This screen enables the designer to select the numeric variable for data mining analysis. This screen contains the following controls:

1. *Numeric Variable List*

A list of the numeric variables in the selected table is displayed. The designer can select one numeric variable from this table.

2. *Sample Data Mining result Textbox*
This textbox displays a sample data mining result involving the numeric variable selected by the designer. (The sample result displayed changes as and when the numeric variable selected by the designer changes)

3. *Next Button*
Pressing this button causes this screen to close, to be replaced by Screen 11.

4. *Previous Button*
Pressing this button causes this screen to close, to be replaced by Screen 9.

5. *Cancel Button*
Pressing this button causes the Analyzer to exit.

6. *Tutorial Button*
Pressing this button causes the tutorial for this screen to start.

2.6.11 Screen 11 : Selecting Variables of interest for Data Mining

This screen enables the designer to select the variable(s) for data mining. It contains the following controls:

1. *Variable List*
This is a list of the variables in the selected table. The designer can select one or more variables from this list

2. *Sample Data Mining result Textbox*
This textbox displays a sample data mining result involving the variable(s) and numeric variable selected by the designer. The sample result displayed should change as and when the designer makes changes in the variable(s) selected for analysis.

3. *Next Button*
Pressing this button causes this screen to close, to be replaced by Screen 12.

4. *Previous Button*
Pressing this button causes this screen to close, to be replaced by Screen 10.

5. *Cancel Button*
Pressing this button causes the Analyzer to exit.

6. *Tutorial Button*
Pressing this button causes the tutorial for this screen to start

2.6.12 Screen 12 : Selecting a decision variable for Data Mining

This screen enables the designer to select a decision variable for data mining. It contains the following controls:

1. *Decision Variable List*
This list contains potential candidates for the decision variable. The designer can select one variable from this list.

2. *Sample Data Mining result textbox*
This textbox displays a sample data mining result involving the variable(s) selected for data mining, the numeric variable and the decision variable. The sample result displayed should change as and when the designer changes the decision variable selection.

3. *Next Button*
Pressing this button causes this screen to close, to be replaced by Screen 13.

4. *Previous Button*
Pressing this button causes this screen to close, to be replaced by Screen 11.

5. *Cancel Button*
Pressing this button causes the Analyzer to exit.

6. *Tutorial Button*
Pressing this button causes the tutorial for this screen to start.

2.6.13 Screen 13 : Selecting the cutoff for Data Mining

This screen enables the designer to set the cutoff for data analysis. It contains the following controls:

1. *Cutoff textbox*
The designer can enter the desired cutoff in this textbox. The default value is 0.02 (2%)

2. *Next Button*
Pressing this button causes this screen to close, to be replaced by Screen 14.

3. *Previous Button*
Pressing this button causes this screen to close, to be replaced by Screen 12.

4. *Cancel Button*
Pressing this button causes the Analyzer to exit.

5. Tutorial Button
Pressing this button causes the tutorial for this screen to start.

2.6.14 Screen 14 : Analysis Description
This screen enables the designer to enter a title for the data mining analysis she has just programmed (e.g., "Analysis of sales by quarter"). It contains the following controls:

1. Analysis Description Textbox
The designer can enter a description for the data mining analysis in this textbox.

2. Define another data mining question Checkbox
The designer can check this box if she wants to define another data mining question.

3. Next Button
Pressing this button causes this screen to close. If the designer has checked the "Define another data mining question" box, Screen 10 is displayed. Else, Screen 15 is displayed.

4. Previous Button
Pressing this button causes this screen to close, to be replaced by Screen 13.

5. Cancel Button
Pressing this button causes the tutorial for this screen to start.

6. Tutorial Button
Pressing this button causes the tutorial for this screen to start.

2.6.15 Screen 15 : Saving the analysis template
This is the standard system File Save Dialog Box. It enables the designer to save the analysis template for future use, if required.

When this dialog box is closed, Screen 16 is displayed.

2.6.16 Screen 16 : Finish
This screen displays a "Finish" message. It contains the following controls:

1. Finish Button
Pressing this button causes the Analyzer to start processing the raw data. Duing the processing, the Analyzer should display suitable progress messages.

2. Previous Button

Pressing this button causes this screen to close, to be replaced by Screen 14.

3. *Cancel Button*
Pressing this button causes the Analyzer to exit, without processing the data.

2.7 The Analyzer Online Tutorial

2.8 Issues

2.8.1 Portability

2.8.2 Data Storage

2.8.3 Multiple Users

2.8.4 Internationalization

3 The VirtualMiner Viewer

3.1 User Interface

3.2 Data Mining and OLAP query results

3.2.1 Collation

3.2.2 Printing / faxing

3.2.3 E-mail

3.2.4 Publishing to website

3.3 The Viewer Online Tutorial

4 Typical Usage Scenario

Blazer Engine
OLAP Support
AF Support
Categorical attributes only
Cutoff for size of dataset
N-P Complete problem (cliques)
HTML